Group of Pictures Effects on Proposed Multiview Video Coding Scheme

Seif Allah El Mesloul Nasri†,*, Abdul Hamid Sadka†, Noureddine Doghmane* and Khaled Khelil‡
†Department of Electronic and Computer Engineering, Brunel University, London, United Kingdom
*Department of Electronics, University Badji Mokhtar, Annaba, Algeria
‡LEER Lab, University Mohamed-Cherif Messaadia, Souk Ahras, Algeria

***Abstract*—In this paper, an investigation of multiview video coding schemes is presented, based on different inter-view prediction structures and various group of pictures (GOP) sizes. The studied inter-view prediction schemes in this paper are: A new approach namely "PIP", recently proposed (PBI), and the benchmark structure (MVC). Results of the conducted tests allow to order the studied inter-view structures and GOP sizes preferences according to their effects in terms of random access and compression efficiency. "PIP" approach achieves significant results in terms of random accessibility yielding a gain of 53.33 % and 36.36 % compared to MVC and PBI approach, respectively. A substantial bit-rate saving is produced by "PBI" approach structure compared to the aforementioned structures. Furthermore, the results ascertain the fact that using a reduced GOP length provides better random access ability and less bit-rate saving. Conversely, larger GOP length leads to low-delay random access and more bit-rate saving.**

***Keywords*—bit-rate; multiview video coding; random access.**

I. Introduction

The Multiview video is based on capturing the same scene from different angles by the use of multiple cameras. This technique offers three-dimension sensation to the viewer, in addition to the possible interaction and free navigation within the displayed scene [1]. Multiview video systems can be applied in many fields such as surveillance, education, video conferencing and telepresence, gaming and 3D cinema [2]. Since the Multiview video sequences contain an important amount of similar data between the views recorded by the adjacent cameras, an exploitation of the inter-view dependencies beside the temporal prediction is required for an efficient Multiview video coding (MVC) process [3][4].

Recent video coding standards such as H.264 [5] and H.265 [6], provide extension profiles allowing the exploitation of the inter-view resemblances for a better compression efficiency.

Further, the compression efficiency and low-delay random access ability come at the top of any video coding standard requirements list [7]. These two requirements, are mainly affected by the applied inter-view prediction scheme and the group of pictures (GOP) length. An inter-view prediction scheme could be described by its combination of view types and views dependency structure [8]. The GOP length describes the hierarchical level of the temporal prediction structure.

In this manuscript, a new inter-view prediction is proposed to improve the random access performance relative to the MVC standard structure and a previous proposed structure (PBI). In addition, an analysis of the effects of changing the GOP length on the employed structures is presented in term of compression efficiency and random access delay.

The remainder of this paper is organized as follows. Section 2 describes some aspects of the multiview video coding and presents the reference inter-view coding approach and a recent related approach. In Section 3, the new proposed prediction structure is introduced. The experimental results including the random access assessment, the GOP sizes effects investigation and the compression efficiency evaluation, are presented and discussed in section 4. Finally, conclusions are provided in section 5.

II. Background

A. *The multiview correlation types*

The multiview video is characterized by three types of correlation [9]. Two types are common with the conventional 2D video, which are the intra picture redundancy and the temporal redundancy. Fig 1 illustrates the hierarchical B algorithm exploiting the temporal redundancy through four GOP sizes.

In addition, the multiview video includes the disparity estimation in the inter-view level to exploit the redundant information between adjacent views.

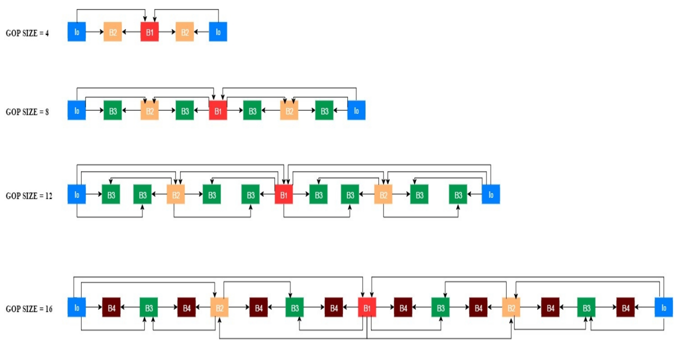

Fig. 1. The temporal prediction architectures of different GOP sizes

Generally, the MVC coding schemes involve the use of 3 types of views. I-view, P-view and B-view. The I-view represents the base view in the MVC scheme; it is used as a reference view for the other types of view. There is a single reference view in the typical inter-view prediction structures. The key and non-key pictures within the group of pictures (GOP) of an I-view are coded without the use of any inter-view reference. The key and non-key pictures of the P-view are

This publication was made possible by NPRP grant 9-181-1-036 from the Qatar National Research Fund (a member of Qatar Foundation).
The statements made herein are solely the responsibility of the authors.

predicted from a unique direction. However, the B-view involves bidirectional inter-view prediction for coding its set of frames.

B. The standard approach IBP

The IBP prediction scheme proposed in [8] which is defined as the default structure of MVC standard uses only a single I-view as a reference view and employs both P and B views. The IBP scheme exploits the inter-view correlation which allows a considerable gain in bit rate and video quality compared to the simulcast scheme [8]. Also, the IBP structure ensures an improvement in the random access performance compared to the IPP structure [8]. The random access ability can be measured according to [10] by the maximum number of reference images N_{max}. N_{max} for IBP is calculated by the following equation:

$$N_{max} = 3 \times H_{max} + 2 + 5 \times [Nbr_{view} - 1] \quad (1)$$

Where H_{max} is the hierarchical level which depends on the GOP size. For example, H_{max}=4 for GOP=8. Nbrview is equal to the number of used view during the coding process. Commonly, IBP prediction scheme in MVC is employed as a benchmark model for performance comparisons between the proposed prediction schemes.

C. The previous proposed approach (PBI)

Unlike the IBP structure, the proposed structure in [11], referred as Proposed 1 or PBI, uses two reference I-views with selected positions which provide a significant acceleration for the random access performance. It also employs the use of more B-views which improve the bit rate gain. The formula that describes the computation of the N_{max} for the PBI structure appears less complicated, as follows:

$$N_{max} = 3 \times H_{max} + 2 \quad (2)$$

Other research [12][13][14] have proposed different MVC structures to improve the random access. However, more improvements need to be done on the random access ability for ensuring better interactivity in the multiview video system.

III. THE NEW PROPOSED STRUCTURE (PIP)

It is known that the type and the views positions within the inter-view structure affect directly the random access performance and the compression efficiency of the MVC codec. The proposed inter-view structure is composed of two base views (I) and six predicted views (P) per group of group of pictures (GGOP).

The two I-views (S2, S5) are independently coded by the use of only the temporal prediction based on the B hierarchical algorithm. The six P-views are coded by a combination of temporal and inter-view predictions, where the maximum hierarchical level equals 3 for GOP=8. S2 and S5 are selected as optimal positions for the reference views; they allow a direct inter-view prediction for all the remaining P views (S0, S1, S3, S4, S6, and S7) without any intermediate view. With this composition, we obtain two sub GGOPs which are completely independent of each other. Every sub GGOP is constructed around an I-view. The first sub GGOP is composed of S2 as a reference view and S0, S1 and S3 as P views. The second sub GGOP is identical to the first one. The key frames of the P views are coded through the key frame of the I view, whereas the non-key frames of the P views are coded from three reference frames, one from the inter-view level of the reference I view and two from their temporal level.

The PIP scheme excludes the use of B views which involve more hierarchical levels that slow down the random access. On the other hand, the key frames of a B-view are coded with two 2 key reference frames and the non key frames are coded with four frames. The employment of two reference I views in addition to the P-views instead of the B-views will significantly improve the random access ability, which means reducing the number of the needed frames for coding or decoding any frame in the multiview video coding structure. This leads directly to reducing the encoding time duration.

Fig 2 depicts the prediction scheme of the PIP structure. The MVC, PBI and PIP structures share the same temporal prediction algorithm. The differences appear in the inter-view prediction schemes, clearly depicted in Fig 3. For the case of 8 views and GOP size=8, the maximum hierarchical level for the "PIP" structure PIP is equal to 4, it is represented in Fig 2 by the green colour (B3). This frame can be found within both I and P views. N_{max} represents the maximum number of reference images needed for coding or decoding the highest hierarchical level frame. The B3 frames of the P views have the highest hierarchical level. To calculate N_{max}, we take the example of B3 frame located in S0/T1 (Fig 2). Accessing this frame requires four reference frames in the temporal level with these positions: S0/T0, S0/T2, S0/T4 and S0/T8, and five reference frames in the inter-view level with the following positions: S1/T0, S1/T1, S1/T2, S1/T4 and S1/T8. Consequently, the equation that describes the calculation of N_{max} for the PIP scheme can be deduced as:

$$N_{max} = 2 \times H_{max} + 1 \quad (3)$$

Where H_{max} is equal to 4 for GOP size = 8. Table 1 shows that the PIP structure provides the simplest calculation of N_{max} compared to the reported structures.

TABLE 1. N_{MAX} EQUATION COMPARISON

	N_{max} equations
MVC	$N_{max} = 3 \times H_{max} + 2 + 5 \times [Nbr_{view} - 1]$
PBI	$N_{max} = 3 \times H_{max} + 2$
PIP	$N_{max} = 2 \times H_{max} + 1$

The following equations calculate de number of necessary reference frames to be crossed for accessing any given picture in the PIP structure:

- For the key frames:

  $$Nbr_{view} = \{0 \text{ for I frames, } 1 \text{ For P frames}\} \quad (4)$$

- For the non-key frames:

  $$Nbr_{view} = \alpha \times H_{level} + \beta \quad (5)$$

$$Where \begin{cases} \alpha = 1,\ \beta = 0\ For\ I\ nonkey\ frames \\ \alpha = 2,\ \beta = 1\ For\ I\ nonkey\ frames \end{cases}$$

H_{level} takes the value of the hierarchical level of the frame in GOP. The equations (4) and (5) demonstrate the simplicity of the calculations required for computing random access in the "PIP" approach.

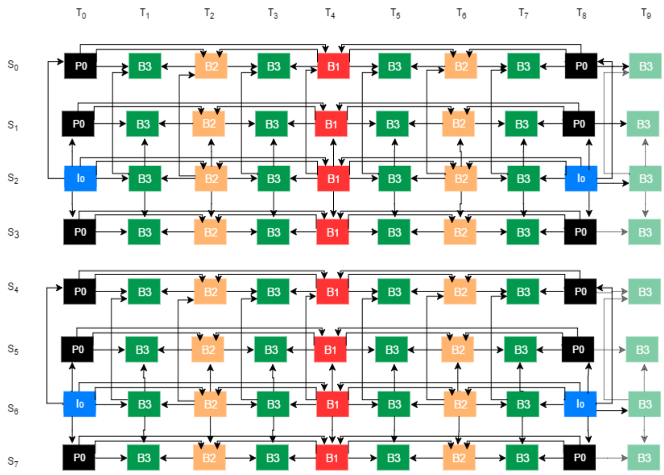

Fig. 2. "PIP" prediction structure (Proposed 2)

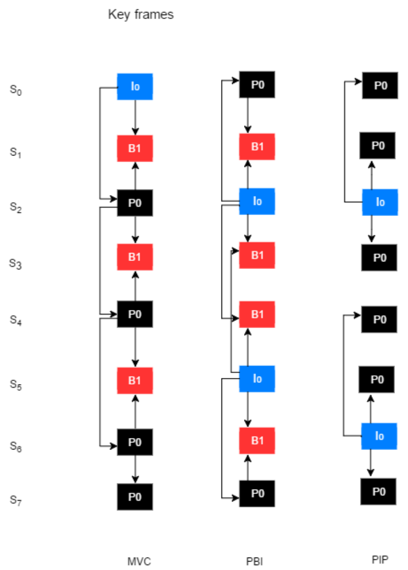

Fig. 3. Key Frames comparison among MVC, PBI and PIP structures

IV. Experimental Results

A. Random access evaluation

In this section, we evaluate the random access performance of the "PIP" coding scheme (PIP) with respect to PBI and MVC schemes. In addition, we investigate the effects of changing the GOP size over the random access performance.

Two metrics are used for the evaluation, the Nmax previously discussed and the global random access evaluation [11], which allows for a full assessment of the considered structure by taking into account the random access cost of each existing picture in the MVC scheme. The random access ability evaluation GR consists of measuring the average cost of encoding all the frames in the GGOP of the structure. The GR parameter is given by the following equation:

$$G_R = \frac{\sum_{i=1}^{V_n}\sum_{t=1}^{GOP(size)}[Nbr_{img}(i,t)]}{GGOP(size)} \qquad (6)$$

Where:

Nbrimg is the number of the decoded frames to access a selected picture in the structure. i indicates the frame position in the view level. t indicates the frame position in the temporal level. Vn is the number of the views in the structure. *GGOP(size)* is the group of the group of picture, and is equal to V_n multiplied by the size of the group of pictures, GOP.

The GOP (size) takes four values in our experiments, which are 4,8,12 and 16. The results of the application of (6) following the GOP size values for each structure are regrouped in table 2.

The calculation method of the random access gain according to GR is defined in (7).

$$\Delta G_R = \frac{G_R(\text{structure 1}) - G_R(\text{structure 2})}{G_R(\text{strcuture 1})} \times 100\% \qquad (7)$$

For instance, for calculating ΔGR (PIP/IBP), GR (structure 1) takes the value of GR (IBP) and GR (structure 2) takes the value of GR (PIP). The rest of examples is revealed in Fig 4.

TABLE 2. G_R VALUES THROUGH DIFFERENT GOP SIZES

GOP	G_R (IBP)	G_R (PBI)	G_R (PIP)
4	7.84	5.75	4.25
8	7.84	6.78	6
12	10.055	8.75	6.58
16	11.3125	10.25	7.75

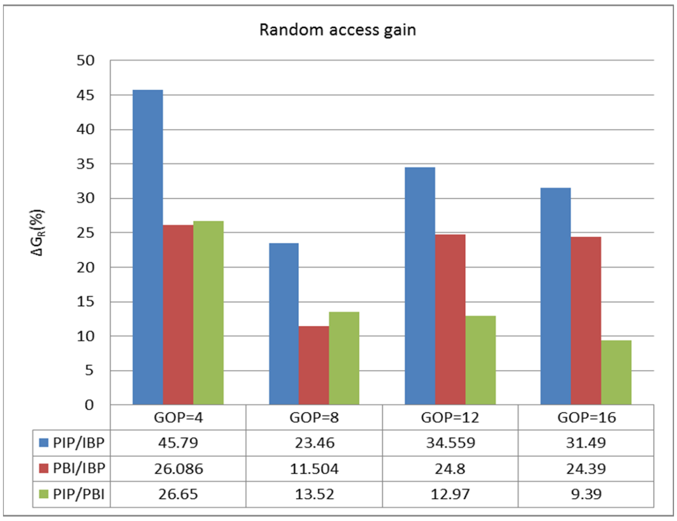

Fig. 4. ΔG_R (%) comparison through different GOP sizes

Table 2 reports the global random access values of the MVC (IBP), Proposed 1 (PBI) and Proposed 2 (PIP) structures, for GOP size=[4,8,12,16]. Additionally, Fig 4 shows the random access gain comparison between the three reported structures, over the four GOPs values.

Overall, it can be inferred clearly from table 2 that the PIP structure gives the lowest GR values. This fact is illustrated in Fig 4 by the best gains that PIP provides compared to the MVC and PBI schemes.

Initially, when the GOP length is equal to 4, the GR attains the lowest values for all structures, where GR (IBP) = 7.84, GR (PBI) =5.57 and GR (PIP) = 4.45. However, the highest GR values are obtained when the GOP length is equal the 16. Moreover, the GR values are ascended generally with the increase of the GOP length from 4 to 16. This is due to the temporal reference which varies depending on the GOP lengths.

Fig 4 shows a comparison, in terms of the random access gain, between MVC, PBI and PIP schemes. It can be clearly revealed that the "PIP" scheme is more effective, with an average gain of ~ 34 % and ~ 16 % relative to MVC and PBI, receptively. Note that the largest gains are achieved when GOPs=4. It exceeds 45 % and 26 % compared to MVC and PBI, respectively.

TABLE 3. G_R VALUES THROUGH DIFFERENT GOP SIZES

GOP	*N_{max}(IBP)*	*N_{max} (PBI)*	*N_{max}(PIP)*
4	15	11	7
8	18	14	9
12	18	14	9
16	21	17	11

Table 3 regroups the N_{max} results of the considered structures. The obtained values demonstrate that the PIP structure reduces significantly the maximum number of the reference frames required for decoding a given frame. Fig 5 expresses this reduction by the improvements in random access ability achieved by the PIP structure, where the maximum N_{max} gain exceeds 53 % and 26 % relative to IBP and PBI structures, respectively. Both G_R and N_{max} results demonstrate that the PIP structure reduces considerably the complexity for accessing a given frame in the multiview video coding scheme, which leads in turn to an enhanced random access ability of the PIP multiview video coding scheme.

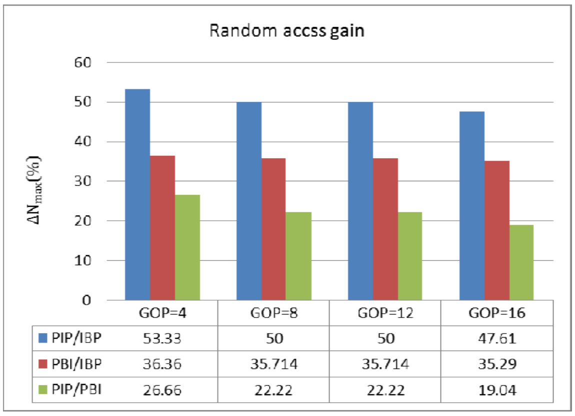

Fig. 5. Δ N_{max} (%) comparison through different GOP sizes

B. *Compression efficiency evaluation*

The compression efficiency experimental results of the studied multiview coding approaches are provided and discussed in this section. The compression efficiency is expressed on graphs of peak signal-to-noise ratio PSNR (dB) versus bit rate (kbit/s). Common initial conditions and data sets have been used in order to provide fair comparisons. Table4 regroups the used multiview video sequences and the common encoding configuration.

TABLE 4. DATA SET AND ENCODING CONFIGURATION

Parameter	**setting**
Video sequences	Race1, Exit and Ballroom
GOP size	4, 8, 12 and 16
Quantization parameter	20, 23, 26 and 29
Symbol mode	CABAC
Search mode	Fast search
Search range	64

The experimental tests about the GOP sizes effects on the Multiview video coding are conducted using three different video sequences (Race1, Exit and Ballroom).Each used video sequences is composed of 8 parallel views. Furthermore, all the sequences share the same resolutions of 640×480. The frame rate for both Exit and Ballroom sequences is equal to 30 fps, and 25 fps for Race1 sequence.

Four GOP sizes have been used, which are identical to the ones used for the random access evaluation tests [4, 8, 12, and 16].The quantization parameter (QP) values control the compression efficiency variations; the lower the value of the QP, the higher the bit rate and video quality.

The symbol mode specifies the used entropy coding mode; Context-adaptive binary arithmetic coding (CABAC) usually enhances the coding efficiency. The fast motion search algorithm, with a search range of 64, is employed since it reduces significantly the encoding time.

Four QP values are used for the three considered multiview coding structures. In addition, four GOP sizes are employed for each QP value, which results in a total of 48 experimental tests.

It is clearly inferred from these results that the employment of a smaller GOP size, regardless of the inter-view prediction structure, always results in a better video quality compared to a longer GOP.

This quality degrades each time the GOP size increases. However, a reduced GOP size generates larger bit rate values. Fig 6, 7 and 8 accurately clarify this point. The obvious example is when the GOP size is equal to 4; the video quality for this case is improved by around 1dB compared to the quality provided by the rest of GOPs. At the same time, the resulting bit rate values are distinctly increased by approximately 500 kbps.

A shorter GOP size leads to a reduction in the distance between the temporal reference pictures, which results in higher level of resemblance between the pictures within the GOP. Subsequently, both the video quality and bite rate values are increased.

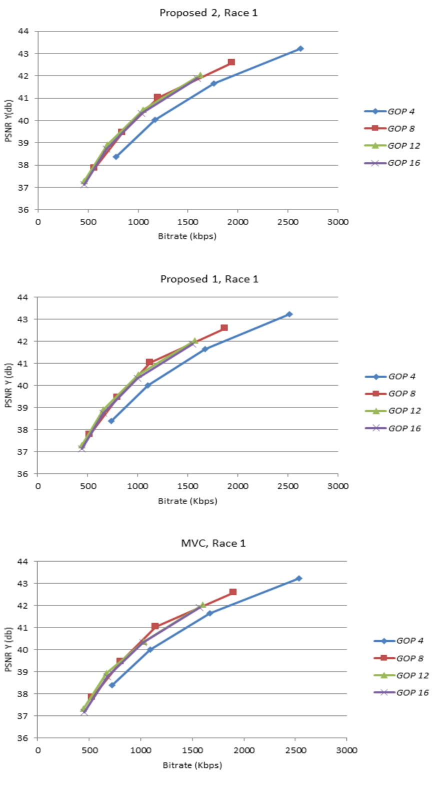

Fig. 6. Compression efficiency evaluation through different GOPs using Race1 sequence

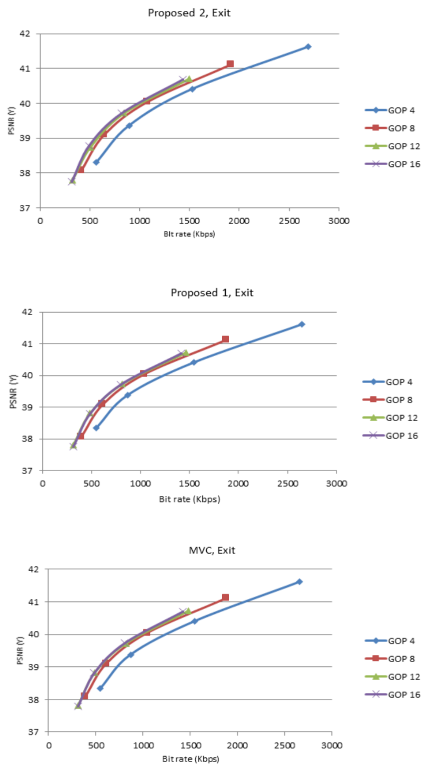

Fig. 7. Compression efficiency evaluation through different GOPs using Exit sequence

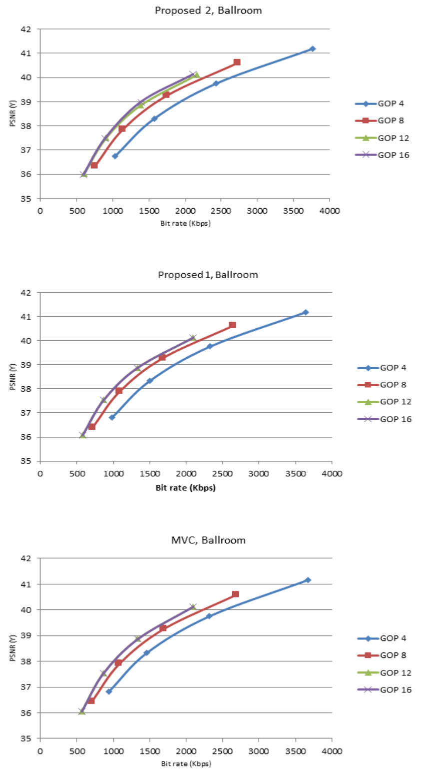

Fig. 8. Compression efficiency evaluation through different GOPs using Ballroom sequence

Fig 9, 10 and 11 depict the compression efficiency results of the two proposed structures compared to the benchmark structure MVC. Only the highest and lowest QP values are selected (QP=20, 29) in order to highlight the comparison in its higher and lower borders. The Δ bit rate illustrated in Fig 9, 10 and 11 has been calculated using the following equation:

$$\Delta \text{bit rate} = \frac{\text{bit rate(MVC)} - \text{bit rate(proposed)}}{\text{bit rate(MVC)}} \times 100\% \qquad (8)$$

Where bit rate (proposed) can take the value of (PBI) or (PIP). The positive results of Δ bit rate will be taken as a gain. However, the negative ones will be considered as a loss.

In general, Fig 9, 10 and 11 illustrate that the Δ bit rate of both PBI and PIP structures increases consistently with the larger GOPs.

For the highest video quality represented by QP=20, the "PBI" scheme constantly provides a positive bit rate gain compared to MVC scheme. However, the PIP scheme delivers a less efficient compression in term of bit rate compared to MVC scheme. Nevertheless, this difference is reduced by the increase in GOPs.

For QP=29 that represents the lowest video quality, an alternation of gain and loss is noted between the PBI and MVC schemes in term of bit rate, whereas "PIP" performs better than MVC in Race 1 when GOPs is equal to 8, 12 and 16. Additionally, PBI gives slightly better results when GOPs is equal to 12 and 16 in Ballroom. However, the MVC scheme provides better results for the rest of cases. Furthermore, it is noted that "PIP" scheme is less efficient along QP=29 regardless of the GOP length.

The composition of the structure is one of the main sources that create the bit rate gain or loss differences.

Including more I-views and P-views will obviously produce additional data during the compression process. This point has been clearly noticed through the PIP structure, which uses two I-views and six P-views. Additionally, the GOP length has a direct effect on the video compression, where the similarity between the successive frames is exploited by the temporal prediction, which is based on the reference pictures that define the start and the end of a GOP. Hence, the similarity considered for exploitation increases as much as the GOP size is enlarged.

Therefore, further data is available to be removed, which leads to improvements in the compression efficiency. However, the resemblance between the frames will gradually decrease along with time. Therefore, it will not be useful to increase the GOP length without any limitation. This fact has been confirmed from the results of Fig 6, 7 and 8, especially when GOP= [12, 16] where practically similar graph lines are clearly noted.

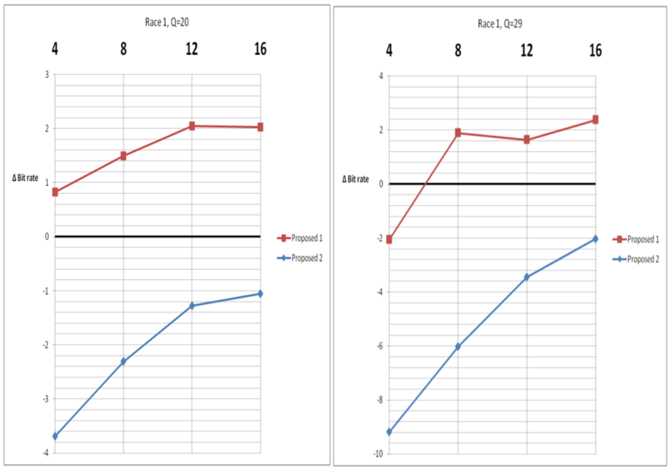

Fig. 9. Δ bit rate of the proposed approaches relative to standard MVC using Race1 sequence

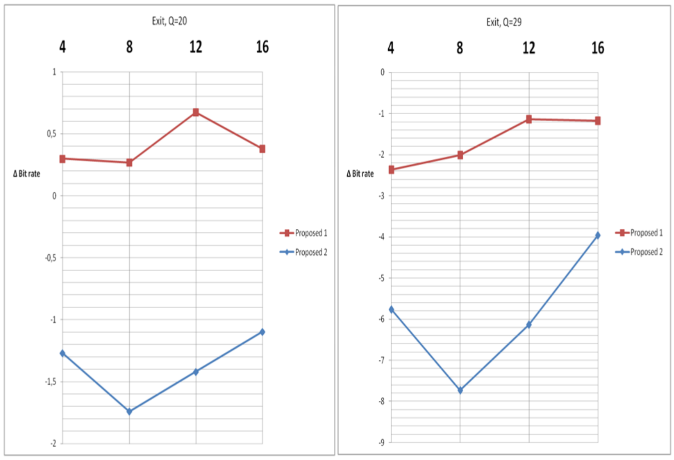

Fig. 10. Δ bit rate of the proposed approaches relative to standard MVC using Exit sequence

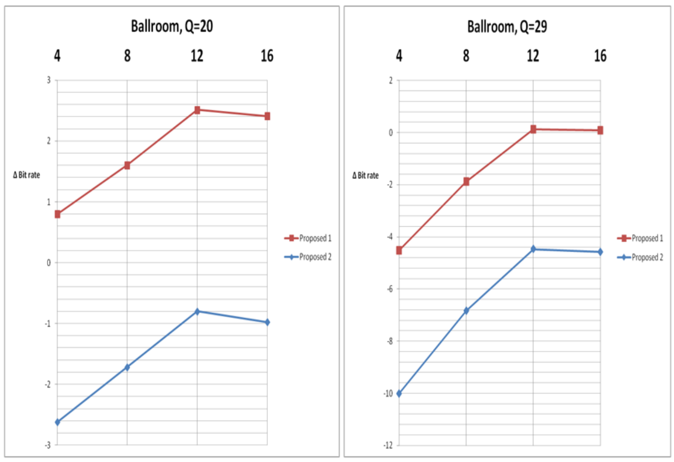

Fig. 11. Δ bit rate of the proposed approaches relative to standard MVC using Ballroom sequence

C. *Prediction structures Trade-offs*

In this section, 3D graphs are presented to highlight the proposed structures trade-off between the GOP size, the bit rate gain and the random access ability. Each 3D graph provides a holistic overview of the experimental outcomes, by illustrating he previous results in one graph composed of 3 axes. Where, the (x) axis shows the four used GOP sizes; the (y) axis represents the average results of the random access ability relative to the MVC structures.

The Δbit rate results, which represent here the average value of the four used quantization parameters (QP=20, 23, 26 and 29), are projected along the vertical axis (z). These 3D charts show the evolutionary effect of changing the GOP size on both the random access ability and the compression efficiency in terms of bit rate.

It can be clearly inferred from the charts that the PBI structure performs better than the PIP structure in term of bit rate saving. Conversely, the PIP outperforms PBI in its random access ability, where the optimum random access efficiency for PIP is achieved when GOP size is equal to 4. Hence, PIP is suitable for the multiview video and Free viewpoint video applications when the bit rate saving is less important and smoother interactivity is required. The PBI can be considered as a balanced structure for standard applications as it provides improvement in both compression efficiency and random access ability relative to the benchmark coder MVC.

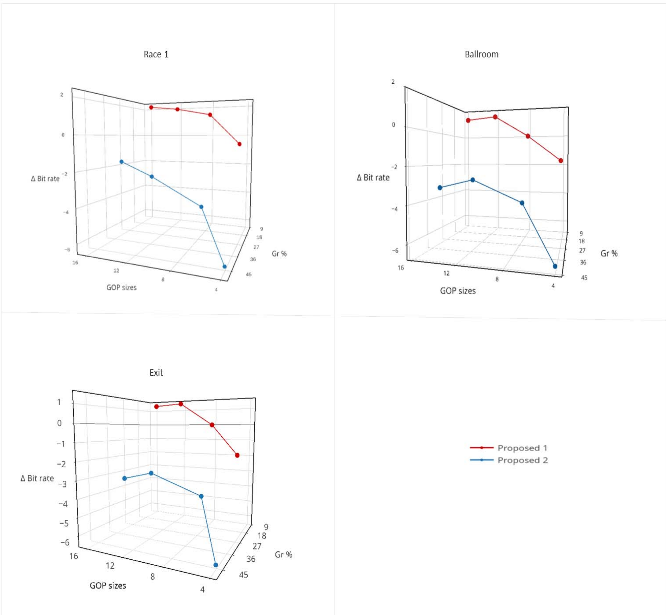

Fig. 12. 3D trade-offs for PBI and PIP structures

V. Conclusion

In this paper, a novel inter-view prediction structure is presented. The so-called PIP or Proposed 2 structure improves the random access ability for the multiview video coder. Mainly composed of two reference views (S2, S6) and a remaining set of P views, the PIP structure is designed to allow faster and direct inter-view prediction scheme. Four GOPs are applied for comparing and evaluating the random access ability and the compression efficiency of the reported multiview coding approaches. The evaluation results have been divided into three parts. Firstly, the random access performance is assessed using two metrics, namely N_{max} and G_R. The PIP coding structure achieves significant G_R gains that exceeds 45 % and 26 % compared to MVC and PBI structures, respectively. Secondly, the compression efficiency is evaluated in terms of bit rate saving and video quality using four GOPs, where for almost all cases the PBI approach achieves the best efficiency. The MVC structure comes in the second place, while the PIP approach shows a remarkable loss in bit rate saving. Finally, a 3D trade-off illustration is presented, showing the GOP size effects on the compression efficiency and random access ability for PBI and PIP coding structures relative to MVC.

References

[1] Philip Benzie et al, " A Survey of 3DTV Displays: Techniques and Technologies" , IEEE Trans. Circuits Syst. Video Technol. 17(11), 1647 – 1658 (2007)

[2] A. Smolic et al, "3D video and free viewpoint video—technologies, applications and MPEG standards", presented at IEEE Int. Conf. Multimedia and Expo, pp. 2161–2164, IEEE, Toronto, Ontario, Canada (2006).

[3] A. Smolic, P. Kauff, "Interactive 3-D video representation and coding technologies", Proc. IEEE Special Issue on Advances in Video Coding and Delivery, vol. 93, no. 1, pp. 98-110, Jan. 2005.

[4] A. Smolic, P. Kauff, "Coding Algorithms for 3DTV—A Survey", IEEE Trans. Circuits Syst. Video Technol. . 17(11), 1606 – 1621 (2007)

[5] J. Ostermann et al, " Video coding with H.264/AVC: tools, performance, and complexity ", IEEE Circuits and Systems Magazine. 4(1), 7 – 28 (2004).

[6] G. J. Sullivan et al, "Overview of the High Efficiency Video Coding (HEVC) Standard ", IEEE Circuits and Systems Magazine. 22(12), 1649 - 1668 (2012).

[7] “Requirements on multi-view video coding v.4,” ISO/IEC JTC1/SC29/WG11, Doc. N7282, Poznan, Poland (2005).

[8] P. Merkle et al, “Efficient prediction structures for multiview video coding,” IEEE Trans. Circuits Syst. Video Technol. 17(11), 1461–1473 (2007).

[9] Anthony Vetro et al, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard" Proceedings

[10] U. Fecker and A. Kaup, “Complexity evaluation of random access to coder multi-view video data,” ISO/IEC JTC1/SC29/ WG11, N8019 (2006).

[11] S. Nasri et al, “Enhanced view random access ability for multiview video coding,” J. Electron. Imaging 25(2), 023027 (2016).

[12] A. Bekhouch, and N. Doghmane. “Multiview video coding with an improved prediction structure for faster random access”, Journal of Electronic Imaging 22(4), 043010 (Oct–Dec 2013).

[13] X. Lv, L. Ma, and J. Guo, “Multiview video coding scheme based upon enhanced random access capacity,” Int. J. Comput. Sci. Issues 10(1), 285 (2013).

[14] A. Bekhouch, and N. Doghmane. “Proposing a new evaluation metric to improved view random access for multi-view video coding“. IPTA 2014: 167-172